Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at http://about.jstor.org/participate-jstor/individuals/early-journal-content.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

American Anthropologist

NEW SERIES

Vol. 18 January–March, 1916 No. 1

ON THE VARIETY OF LINES OF DESCENT REPRESENTED IN A POPULATION

By FRANZ BOAS

THE degree of variation of the individuals composing a population is expressed by means of tables showing the distribution of the values of the particular measurement that is being considered; and the range of variation may be studied by the usual methods, particularly by means of the standard deviation of the series.

In this case the individuals are considered as representatives of the whole series, without any regard to the descent of each. It may therefore well be that, from a biological point of view, two series having the same degree of variability of the component individuals are quite distinct. In one extreme case we may imagine the population as consisting of a number of entirely distinct lines of descent, in which each line represents a particular type. Then by grouping together the individuals of each type, we obtain a series of types different among themselves, which would indicate the heterogeneity of the types that enter into the constitution of the whole population. In the other extreme case we may imagine all the lines of descent composing the population to represent the same ancestral strain. Obviously in this case every single ancestral line would be representative of the whole population, and the averages of all the different lines would be the same.

In other words, in the former case, each line of descent will be characterized by certain traits; in the latter case, all lines of descent will be the same. This latter case would be the result of an absolute inbreeding; and there could no longer be any family resemblance, because all the families would be alike, and in every family all the varying types of the population would be represented. A fraternity would be an adequate representation of the whole people or of a local type.

This case, of course, cannot possibly occur in man. It corresponds to the individuals composing the pure lines of Johannsen, which I should prefer to term "lines of identical descent." In man we can expect only an approach to these conditions, that may develop with continued inbreeding, such as is characteristic of stable rural communities or of primitive tribes in thinly settled countries.

The other extreme is represented by new dense populations, like that of the United States, in which there has not been time for a permeation of the constituent elements. The conditions that develop in a mixed white and negro population will illustrate the characteristics of these conditions. If in such a population whites and negroes do not intermingle, the racial position of each individual will at once be indicated by the fact that all the members of his family will possess the same racial traits, or show a marked resemblance in this respect. On the other hand, in a mulatto population, such differences between members of families cannot exist, but all family strains will show equally effects of mixture. In the former case we have a strong selective mating, in the latter case, none.

Thus it appears that strong family resemblances in a population are indications of the heterogeneity of the composite lines of descent.

The heterogeneity of which I speak here is therefore not an expression of the purity of racial descent, for identity of the lines of descent represented by the families of the population may well be found in cases in which a few ancestors belonged to distinct racial types, but in which the modern population developed by long-continued inbreeding without selective mating. It is rather an expression of the differences between the various lines of descent represented in a population.

From all I have said, it appears that, in a population in which the lines of descent are the same for each fraternity, we should expect the averages of these fraternities to be identical, provided the fraternities are large enough. In populations in which the lines of descent are quite distinct, we should expect the averages of the fraternities to differ strongly among themselves. In other words, the variability of the average of the fraternities will be a measure of the heterogeneity of the lines of descent.

This value, however, cannot be obtained by direct observation, because in man each fraternity comprises a few individuals only.

We will call the general average of the measurement studied a; its standard variation, σ; the number of children in a certain group, m; and the deviation for each child from the general average, x. Then the average for each family is

$$a = \frac{x_1 + x_2 + \cdots + x_m}{m},$$

and therefore the variability of these averages.

$$s_m{}^2 = \left[\left(\frac{x_1 + x_2 + \cdots + x_m}{m}\right)^2\right],$$

where the brackets indicate the process of averaging. If we call the coefficient of fraternal correlation r, we have

$$s_m{}^2 = \frac{1 + (m - 1)r}{m}\sigma^2.$$

For a very large number of children,

$$s^2 = r\sigma^2 = [xx].$$

The heterogeneity of descent may therefore be determined directly from the two values r and σ.

In calculating the coefficient of fraternal correlation, it is necessary to treat families with 2, 3, 4, etc., children separately, because the weights of the coefficients of correlation obtained from each group are not the same. This is due to the fact that in fraternal correlations each individual appears a number of times, as he is correlated with each member of the fraternity.

In our series we are dealing with large numbers of cases, so that the value of the coefficient and also the value of σ may be assumed to be known accurately. The coefficient is calculated from the products $x_a x_b$. If f is the number of families, the average product

$$[xx] = \frac{1}{f} \Sigma \frac{x_1x_2 + x_1x_3 + \cdots + x_ax_b + \cdots + x_{m-1}x_m}{\frac{1}{2}m(m-1)}.$$

Its standard error is

$$\epsilon^2 = \left[\left\{ \frac{1}{f} \cdot \Sigma \frac{x_1x_2+x_1x_3+\cdots+x_ax_b+\cdots x_cx_d+\cdots+x_{m-1}x_m}{\frac{1}{2}m(m-1)} - r\sigma^2 \right\}^2 \right].$$

When squaring the sum, the products of the terms representing different families are all independent, and each will be, on the average, $r\sigma^2$. There are $f(f-1)$ products of this type. Therefore

$$\epsilon^2 = \frac{\Sigma x_a^2x_b^2 + \Sigma x_a^2x_bx_c + \Sigma x_ax_bx_cx_d}{f^2\frac{1}{4}m^2(m-1)^2} + \frac{f(f-1)}{f^2} r^2\sigma^4 + r^2\sigma^4$$

$$- 2r\sigma^2 \left[\frac{1}{f} \cdot \Sigma \frac{x_1x_2 + \cdots + x_ax_b + \cdots x_{m-1}x_m}{\frac{1}{2}m(m-1)} \right].$$

The last average is $r\sigma^2$.

There are $f\dfrac{m(m-1)}{1 \cdot 2}$ members of the type $x_a^2x_b^2$

$f\,m(m-1)(m-2)$ " " " " $x_a^2x_bx_c$

$f\dfrac{m(m-1)(m-2)(m-3)}{4}$ " " " " $x_ax_bx_cx_d$

Therefore

$$\epsilon^2 = \frac{2[x_a^2x_b^2] + 4(m-2)[x_a^2x_bx_c] + (m-2)(m-3)[x_ax_bx_cx_d]}{fm(m-1)} - \frac{r^2\sigma^4}{f}$$

$$[x_a^2x_b^2] = r^2[x_a^4] + (1-r^2)\sigma^2[x_a^2]$$

$$= (1+2r^2)\sigma^4$$

$$[x_a^2 x_b x_c] = \frac{r}{1+r}[x_a^2 x_b(x_a + x_b)]$$

$$= \frac{r}{1+r}([x_a^3 x_b] + [x_a^2 x_b^2])$$

$$= \frac{r}{1+r}(r[x_a^4] + (1 + 2r^2)\sigma^4)$$

$$= r(1 + 2r)\sigma^4$$

$$[x_a x_b x_c x_d] = \frac{r}{1+2r}[x_a x_b x_c(x_a + x_b + x_c)]$$

$$= 3r^2\sigma^4$$

By substitution,

$$\epsilon^2 = \frac{2(1+2r^2)+4(m-2)(r+2r^2)+(m-2)(m-3)3r^2-m(m-1)r^2}{fm(m-1)}\sigma^4$$

$$= \frac{1 + 2(m-2)r + (m^2 - 3m + 3)r^2}{f\dfrac{m(m-1)}{2}}\sigma^4 .$$

If we designate the total number of individuals $fm = n$,

$$\epsilon^2 = \frac{2}{m-1} \cdot \frac{1 + 2(m-2)r + (m^2 - 3m + 3)r^2}{n}\sigma^4 .$$

From this we obtain for various families the following values of ϵ^2:

For 2 children	$\frac{2}{1}\frac{1 + r^2}{n}\sigma^4$
“ 3 “	$\frac{2}{2}\frac{1 + 2r + 3r^2}{n}\sigma^4$
“ 4 “	$\frac{2}{3}\frac{1 + 4r + 7r^2}{n}\sigma^4$
“ 5 “	$\frac{2}{4}\frac{1 + 6r + 13r^2}{n}\sigma^4$
“ 6 “	$\frac{2}{5}\frac{1 + 8r + 21r^2}{n}\sigma^4$
“ 7 “	$\frac{2}{6}\frac{1 + 10r + 31r^2}{n}\sigma^4$

For values of r, this gives the following values for $\frac{n\epsilon^2}{\sigma^4}$:

r	0.00	0.10	0.20	0.30	0.32	0.34	0.36	0.38	0.40	0.42	0.44	0.46	0.48	0.50
2 children...	2.00	2.02	2.08	2.18	2.20	2.23	2.26	2.29	2.32	2.35	2.39	2.42	2.46	2.50
3 " ...	1.00	1.23	1.52	1.87	1.95	2.03	2.11	2.19	2.28	2.37	2.46	2.55	2.65	2.75
4 " ...	.67	.98	1.39	1.89	2.00	2.11	2.23	2.35	2.48	2.61	2.76	2.88	3.02	3.17
5 " ...	.50	.86	1.36	1.98	2.12	2.27	2.42	2.58	2.74	2.90	3.08	3.25	3.44	3.62
6 " ...	.40	.80	1.38	2.12	2.28	2.46	2.64	2.83	3.02	3.22	3.44	3.65	3.87	4.10
7 " ...	.33	.77	1.41	2.26	2.46	2.66	2.87	3.09	3.32	3.56	3.80	4.05	4.31	4.58

From these may be derived the following relative weights:

r	0.00	0.10	0.20	0.30	0.32	0.34	0.36	0.38	0.40	0.42	0.44	0.46	0.48	0.50
2 children...	50	49	48	46	45	45	44	44	43	43	42	41	41	40
3 " ...	100	81	66	53	51	49	47	46	44	42	41	39	38	36
4 " ...	150	102	72	53	50	47	45	43	40	38	36	35	33	31
5 " ...	200	116	74	50	47	44	41	39	37	34	32	31	29	28
6 " ...	250	125	72	47	44	41	38	35	33	31	29	27	26	24
7 " ...	300	130	71	44	41	38	35	32	30	28	26	25	23	22

I have computed, according to this method, the correlations for several nationalities. The material is the same as that on which my reports on the "Changes in Bodily Forms of Immigrants" is based.[1] I have computed the coefficients of fraternal correlation for the cephalic index, assuming the same correction for all children which I developed in the book just referred to; namely, a decrease of the cephalic index of 0.14 unit per year until the twentieth year is reached; and a reduction of females to male values by assuming from 17 years on a difference of 0.5 unit, the men being more long-headed than the women.

The actual calculation was made according to the formula

$$\Sigma xx = \tfrac{1}{2}\{(\Sigma x)^2 - \Sigma x^2\}.$$

SQUARES OF STANDARD DEVIATION OF xx.

No. of Children in Family	Bohemian			Central Italian		
	No. of Families	Variation of xx		No. of Families	Variation of xx	
		Observation	Theory		Observation	Theory
2................	253	217	183	193	187	209
3................	94	116	131	108	94	147
4................	52	93	111	62	105	125
5................	16	96	100	30	174	112
6................	5	31	94	17	50	80

---

[1] *Changes in Bodily Form of Descendants of Immigrants* (final report). Columbia University Press, 1912.

I have also calculated for two types the standard variations of the values xx and their averages, in order to give a means of comparison of the theoretical values given above and the empirical values. The variation of xx is of course strongly skew.

Following are the results for a few populations:

	$[xx]$	s
Bohemians	5.60	± 2.37
Central Italians	5.72	± 2.39
(Potenza)	(5.83)	(± 2.41)
East European Jews	5.25	± 2.29*
Worcester, Mass.	5.46	± 2.36
Scotch	4.72	± 2.17

* Calculated from 255 families of 2 children.

It appears from these few samples that the variation in lines of descent in Europe seems to be very uniform. The selection of Potenza, a smaller district of Central Italy, does not show a decrease of the variability of lines of descent.

Theoretically the interesting and important question would be, how low the values of the heterogeneity of lines of descent may be in extreme cases. As indicated before, it seems very likely that, with long-continued inbreeding without selective mating, the average for the families might be very much alike.

I have no material from European villages that would lend itself to a test of this problem; and reliable material from primitive tribes is difficult to obtain, because in most cases the actual biological relationships are uncertain. I am indebted to Professor Eugen Fischer for data on the Bastards of South Africa that were not published in his volume.[1] I have treated them in the same manner as the European material. There are 75 individuals in all.

4 families of 2 children.
4 " " 3 "
3 " " 4 "
2 " " 5 "
3 " " 6 "
1 family " 7 "
1 " " 8 "

---

[1] *Die Rehobother Bastards.* Jena, 1913.

The observations give an average for the whole series of 75.9; $\sigma^2 = 6.5$.

$$\text{Bastards } [xx] = 1.6 \quad s = \pm 1.26$$

In this series, in which descent is probably in the great majority of cases reliable, and in which we have long-continued inbreeding, we have therefore much greater uniformity of the various family averages. The calculation of observations collected in 1892 among the eastern Chippewa of Canada gives similar results. The data were collected among the Missisauga and other bands north of Lake Huron. I have the following family groups at my disposal:

	Missisauga	Other Chippewa
Families of 2 children	38	23
" " 3 "	10	6
" " 4 "	6	4
" " 5 "	2	3
" " 6 "	1	–

There are 146 and 95 individuals respectively. The series give the following results:

Missisauga	$a = 80.4$	$\sigma^2 = 12.0$	$[xx] = 2.40$	$s = \pm 1.47$
Other Chippewa	82.1	14.1	3.14	$s = \pm 1.77$

In both cases we have populations containing many half-breeds, but, on the whole, an old inbred mixture.

The results given here might be doubted, on account of the uncertainty of parentage among Indians. However, this condition is always clearly expressed in the coefficients of correlation obtaining between father and child. There is hardly ever a doubt as to the relation between child and mother, so that this correlation can be taken as nearly correct. I have calculated for both groups of Chippewa the correlations between parents and children. I give the values of $[xy]$ and r, and the number of pairs from which they are derived.

	Father and Children.			Mother and Children.		
	$[xy]$	r		$[xy]$	r	
Missisauga	0.65	0.05	*(30)*	1.23	0.10	*(69)*
Other Chippewa	3.30	0.24	*(66)*	2.50	0.16	*(48)*

The correlations between mother and child are low, and on the average not higher than those of the fathers. Even if we might consider the value as a little too low, we are justified in concluding that the correlations are much lower than in European populations. It is easy to see that the correlations between parents and children will also be the lower, the more uniform the lines composing the population. This point has been made by Johannsen in his "Exakte Erblichkeitslehre."

The square standard variation of individuals composing a family, *i. e.*, of the fraternities, is a correlate of the variability of family lines, and must average $\sigma^2(1 - r)$. In the following table I give the variability of the fraternities s_f, and, for the sake of comparison, repeat the variabilities of the family lines s.

Bohemians	$s_f =$ ± 2.61	$s =$ ± 2.37
Central Italians	± 2.72	± 2.39
(Potenza)	± 2.52	± 2.41
East European Jews	± 2.52	± 2.29
Worcester, Mass.	± 2.36	± 2.36
Scotch	± 2.66	± 2.17
Bastards	± 2.52	± 1.26
Missisauga	± 3.10	± 1.47
Other Chippewa	± 3.32	± 1.77

I presume the high values among the Chippewa are due largely to the great difference in type between the component elements of the mixed population, the Irish and French being long-headed, the Chippewa short-headed. It is possible, however, that inaccuracy of information regarding the actual relationships accounts for part of these phenomena.

COLUMBIA UNIVERSITY,
NEW YORK CITY